ON THE STRONG CONSISTENCY OF APPROXIMATE MAXIMUM LIKELIHOOD ESTIMATORS*

by

Michael D. Perlman

Technical Report No. 129

December 1969

*This research was supported by the National Science Foundation under Grant No. GP-9593.

1. Introduction.

Let $\mathcal{P}$ be a set of distinct probability distributions on a measurable space $(\mathfrak{X}, \mathfrak{A})$. Let $\theta = \theta(P)$ be a mapping of $\mathcal{P}$ onto a Hausdorff topological space Θ which satisfies the first axiom of countability. Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed (i.i.d.) random variables assuming values in $\mathfrak{X}$, each distributed according to P_0, and let $\theta_0 = \theta(P_0)$. The symbol P_0 is also used to denote the product probability measure on the infinite product space of all sequences $(x_1, x_2, \cdots)$, and $_*P_0$ denotes the induced inner measure on this space. Let $w(x, \theta)$ be a real-valued function defined on $\mathfrak{X} \times \Theta$ such that for each fixed θ, $w(\cdot, \theta)$ is measurable, and for $n = 1,2,\ldots$ let

$$w_n(\theta) = w_n(x_1, \ldots, x_n, \theta) \equiv \frac{1}{n} \sum_{i=1}^{n} w(x_i, \theta).$$

(For any other function $y(x, \theta)$, $y_n(\theta)$ is defined in a similar manner.) In this paper we discuss the strong consistency of estimators which are based on maximizing $w_n(\theta)$.

Let $\mathfrak{T}_1$ denote the class of all estimating sequences $\{T_n\} = \{T_n(x_1, \ldots, x_n)\}$ (T_n is a Θ - valued function and is <u>not necessarily measurable</u>) such that for all $P_0 \in \mathcal{P}$,

$$(1.1) \qquad {}_*P_0[\sup_{\Theta} w_n(\theta) = w_n(T_n) \text{ fasln}] = 1,$$

(all suprema in this paper are taken with respect to θ over the indicated subset) where "fasln" abbreviates "for all sufficiently large n" and where, if $\{A_n\}$ is any sequence of sets, $\{A_n$ fasln$\}$ is the set

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k\geq n} A_k.$$

If $\{T_n\}$ is in $\mathfrak{J}_1$, we call it a <u>maximum-w</u> <u>estimator</u> (MWE). For reasons stated later, we shall mainly consider the larger class $\mathfrak{J}_2$, consisting of all estimating sequences such that for all P_0 in $\mathcal{P}$

$$(1.2) \qquad {}_*P_0[H(\sup_{\Theta} w_n(\theta), w_n(T_n)) \to 0] = 1,$$

where $H(a, b) \equiv \begin{cases} a-b & \text{if } a < \infty \\ b^{-1} & \text{if } a = \infty \text{ and } b > 0 \\ 1 & \text{if } a = \infty \text{ and } b \leq 0. \end{cases}$

If $\{T_n\}$ is in $\mathfrak{J}_2$, it is called an <u>approximate</u> <u>maximum-w</u> <u>estimator</u> (AMWE).

<u>Example 1</u>. Suppose that $\theta(P)$ is $1 - 1$ and that each P has a density $f(x, \theta)$ with respect to some measure μ. If $w(x, \theta) = \log f(x, \theta)$, then $\mathfrak{J}_1$ contains all maximum likelihood estimators (MLE) and $\mathfrak{J}_2$ contains all approximate maximum likelihood estimators (AMLE) (in the sense of Wald [9], p. 600, Theorem 2).

<u>Example 2</u>. Let $\mathcal{P}$ denote the set of all distributions on $(-\infty, \infty)$ which possess a unique population median. Let $\Theta = (-\infty, \infty)$ and $\theta(P)$ = median of P. If $w(x, \theta) = -|x - \theta|$, then $\mathfrak{J}_1$ contains all sample medians (recall that the sample median may not be uniquely determined).

For each θ_0 in Θ let $\{V_r\} = \{V_r(\theta_0)\}$ be a decreasing sequence of neighborhoods of θ_0 which form a base for the neighborhood system at θ_0 (so $\cap V_r = \{\theta_0\}$), and let $\Omega_r = \Omega_r(\theta_0) = \Theta - V_r(\theta_0)$.

Then an estimating sequence $\{T_n\}$ is <u>strongly consistent</u> if and only if for all P_0 in $\mathcal{P}$ and $r \geq 1$,

$$(1.3) \qquad {}_*P_0[T_n \notin \Omega_r \text{ fasln}] = 1.$$

Note that if this is satisfied for one such sequence of neighborhoods $\{V_r\}$, it must be satisfied for any other such sequence.

It has been customary in previous papers on this subject (e.g. [5], [10], [6], [7], [9]) to present conditions which imply strong consistency for all $\{T_n\}$ in the larger class $\mathfrak{J}_2$, since $\mathfrak{J}_1$ may be empty (the supremum in (1.1) may not be attained) or since it may be more convenient in practice to use an AMWE. A convenient starting point for this problem is the following easy lemma (the proof is omitted).

<u>Lemma 1.1</u>. A necessary and **sufficient** condition for the strong consistency of every AMWE is that for every P_0 in $\mathcal{P}$ and every $r \geq 1$

$$(1.4) \qquad {}_*P_0[\limsup_{n \to \infty} \{\sup_{\Omega_r} w_n(\theta) - \sup_{V_r} w_n(\theta)\} < 0] = 1.$$

Clearly, (1.4) is unchanged if $w(x, \theta)$ and $w_n(\theta)$ are replaced by

$$u(x, \theta) = u(x, \theta; \theta_0) \equiv w(x, \theta) - w(x, \theta_0)$$

$$u_n(\theta) = u_n(\theta; \theta_0) \equiv w_n(\theta) - w_n(\theta_0),$$

and (1.4) is implied by the stronger condition

$$(1.5) \qquad {}_*P_0[\limsup_{n\to\infty} \sup_{\Omega_r} u_n(\theta; \theta_0) < \liminf_{n\to\infty} \sup_{V_r} u_n(\theta; \theta_0)] = 1.$$

Since $u_n(\theta_0; \theta_0) = 0$ and $\theta_0 \in V_r$, (1.5) in turn is implied by

$$(1.6) \qquad {}_*P_0[\limsup_{n\to\infty} \sup_{\Omega_r} u_n(\theta; \theta_0) < 0] = 1.$$

Under some additional assumptions, (1.4) and (1.6) are in fact equivalent (see remarks after Lemma 2.9).

Previous papers treating strong consistency of MWE's or AMWE's have, in essence, presented conditions which imply (1.6). Wald [9] presented essentially the following condition (here weakened so as not to require measurability of $\sup_{\Omega_r} u_1(\theta; \theta_0)$: for all P_0 and r, there exists a measurable real-valued function $s(x) = s(x; P_0, r)$ defined on $\mathcal{X}$ such that $\sup_{\Omega_r} u_1(\theta; \theta_0) \leq s(x)$ for all x in a set of probability 1 (depending on P_0 and r) and $E_0 s(x) < 0$ (see Proposition 2.11). It is easy, however, to find simple examples where this condition fails but all AMWE's are strongly consistent ([6], p. 904). Therefore, several other authors have suggested weaker conditions which imply (1.6). Before discussing these, we introduce several definitions.

Let Γ be a subset of Θ and let $y(x, \theta)$ be a real-valued function, defined on $\mathcal{X} \times \Gamma$, which is measurable in x for each fixed θ. Let the sequence of $\mathcal{X}$- valued i.i.d. random variables $X_1, X_2, \ldots$ have probability distribution P (which need not be in $\mathcal{P}$). (Later we shall take $\Gamma = \Omega_r$, $y(x, \theta) = u(x, \theta; \theta_0)$, and $P = P_0$).

Definition 1. $y(x, \theta)$ is dominated on Γ (with respect to P) if there is a positive integer k and a real-valued function $s(x_1, \ldots x_k)$ on $\mathcal{X} \times \cdots \times \mathcal{X}$, measurable with respect to the product σ - field $\mathcal{G} \times \cdots \times \mathcal{G}$, such that

(i) $\sup_\Gamma y_k(\theta) \leq s(x_1, \ldots, x_k)$ for all $x_1, \ldots, x_k$ in a set of probability one, and

(ii) $E s(X_1, \ldots, X_k) < \infty$.

<u>Remark</u>: Note that if $\sup_\Gamma y_k(\theta)$ is measurable, it can be used in place of $s(x_1, \ldots, x_k)$. In any case, note that s can be chosen to be a symmetric function of $x_1, \ldots, x_k$, for we may replace s by

$$s^*(x_1, \ldots, x_k) = \frac{1}{k!}\sum s(x_{i(1)}, \ldots, x_{i(k)})$$

where the sum is taken over all permutations of $(1, \ldots, k)$. Also, s can be chosen such that $Es(X_1, \ldots, X_k) > -\infty$ for, if not, replaces s by $\max(s, M)$ for any number M.

<u>Definition 2</u>. $y(x, \theta)$ is <u>weakly dominated on</u> Γ (with respect to P) if there exists a function $b(\theta)$ defined on Γ, $0 < b(\theta) < \infty$, such that $y(x, \theta)/b(\theta)$ is dominated on Γ (P).

Replacing $b(\theta)$ by $\max(b(\theta), 1)$ if necessary, we can assume that $\inf_\Gamma b(\theta) > 0$.

<u>Definition 3</u>. $y(x, \theta)$ is <u>dominated by</u> 0 <u>on</u> Γ (P) if Definition 1 holds with (ii) replaced by

(ii)' $Es(X_1, \ldots, X_k) < 0$.

The remark following Definition 1 applies here as well (now M must be chosen so that $E \max(s, M) < 0$).

<u>Definition 4</u>. $y(x, \theta)$ is <u>weakly dominated by</u> 0 <u>on</u> Γ (P) if there exists a function $b(\theta)$ on Γ, $0 < b(\theta) < \infty$, such that $y(x, \theta)/b(\theta)$ is dominated by 0 on Γ (P) and

(iii) $\inf_\Gamma b(\theta) > 0$.

Note that if $b'(\theta)$ $(0 < b'(\theta) < \infty)$ is such that $y(x, \theta)/b'(\theta)$ is dominated by 0 on Γ, it does not necessarily follow that $y(x, \theta)$ is weakly dominated by 0 on Γ, since replacing $b'(\theta)$ by $b(\theta) = \max(b'(\theta), a)$ $(a > 0)$ will not necessarily preserve (ii).

<u>Definition 5</u>. Let $B(\Gamma)$ denote the Banach space of all bounded real-valued functions on Γ (with the usual supremum norm). $y(x, \theta)$ is <u>Bochner-dominated on</u> Γ (P) if there is a positive integer j and a function $v(\theta) = v(x_1, \ldots, x_j, \theta)$ mapping $\mathcal{X} \times \cdots \times \mathcal{X}$ into $B(\Gamma)$ such that

(iv) v is a strongly measurable mapping (with respect to the product σ-field $\mathcal{A} \times \cdots \times \mathcal{A}$)

(v) $\| v \| = \sup_\Gamma |v(\theta)|$ is integrable

(vi) for all $x_1, \ldots, x_j$ in a set of probability one, $y_j(\theta) \leq v(\theta)$ for all θ in Γ simultaneously.

(The reader is referred to [4] for definitions of these terms. See also [3] and [7].)

Note that (iv) and (v) together are equivalent to Bochner integrability of v. Also, v can be chosen to be a symmetric function of $x_1, \ldots, x_j$.

<u>Definition 6</u>. $y(x, \theta)$ is <u>weakly Bochner-dominated on</u> Γ (P) if there is a function $b(\theta)$, $0 < b(\theta) < \infty$, such that $y(x, \theta)/b(\theta)$ is Bochner-dominated on Γ (P).

Again, $b(\theta)$ can be chosen so that $\inf_\Gamma b(\theta) > 0$.

<u>Definition 7</u>. $y(x, \theta)$ is <u>Bochner-dominated by</u> 0 <u>on</u> Γ (P) if

Definition 5 holds and in addition

(vii) $\sup_\Gamma Ev(\theta) < 0$.

Note that (v) of Definition 5 implies that $-\infty < \sup_\Gamma Ev(\theta)$.

Definition 8. $y(x, \theta)$ is weakly Bochner-dominated by 0 on Γ (P) if there exists a function $b(\theta)$, $0 < b(\theta) < \infty$, such that $y(x, \theta)/b(\theta)$ is Bochner-dominated by 0 on Γ (P) and $b(\theta)$ satisfies (iii). (The comment after Definition 4 also applies here).

Clearly, the following implications hold among Definitions 1-8: $3 \Rightarrow 4 \Rightarrow 2$, $3 \Rightarrow 1 \Rightarrow 2$, $7 \Rightarrow 8 \Rightarrow 6$, $7 \Rightarrow 5 \Rightarrow 6$. It will be shown later that $1 \Leftrightarrow 5$, $2 \Leftrightarrow 6$, $3 \Leftrightarrow 7$, $4 \Leftrightarrow 8$.

We now present four conditions, weaker than Wald's condition, which imply (1.6) (as shown in section 2) and therefore imply the strongly consistency of all AMWE's. The first is based on an idea of Berk [2] (probably inspired by Kiefer and Wolfowitz [6]) which is weakened here to avoid topological and measurability restrictions:

Condition B. For every P_0 in $\wp$ and every integer $r \geq 1$, $u(x, \theta; \theta_0)$ is dominated by 0 on $\Omega_r (P_0)$.

Huber's assumptions A2-5 [5] suggest the following condition, again stated in generalized form:

Condition B (weak). For every P_0 and r, $u(x, \theta; \theta_0)$ is weakly dominated by 0 on $\Omega_r(P_0)$.

(In Proposition 2.11 it is shown that Huber's assumptions A2-5 imply B(weak).) The following condition is an extension of a condition introduced by LeCam ([7], p. 303-4):

<u>Condition L</u>. For every P_0 and r, $u(x, \theta; \theta_0)$ is Bochner-dominated by 0 on $\Omega_r (P_0)$.

This in turn can be extended as follows:

<u>Condition L (weak)</u>. For every P_0 and r, $u(x, \theta; \theta_0)$ is weakly Bochner-dominated by 0 on $\Omega_r (P_0)$.

Clearly, $B \Rightarrow B(\text{weak})$ and $L \Rightarrow L(\text{weak})$. Examples are presented in section 2 which show that the reverse implications fail. Also, it is shown that $B \Leftrightarrow L$ and $B(\text{weak}) \Leftrightarrow L(\text{weak})$, and each of these implies (1.6) (for all P_0, r). Under additional restrictions it is shown (Theorem 2.8) that $B \Leftrightarrow (1.6)$ for all P_0, r; however these restrictions are rather strong, so in many applications B(weak) must be verified (see Examples 3 and 4).

In applying B or B(weak), however, it is often difficult to find the function $b(\theta)$ and/or integer k required in Definitions 1-4 (see Examples 4, 5, and 6, and the comments at the end of Section 2). We present a new condition in section 3 which (under mild restrictions) implies (1.6) for all P_0, r and which, since it does not entail the difficulties just mentioned, is often easier to verify. This condition does not require dominance or weak dominance.

Returning to the class $\mathfrak{J}_1$, it is clear that (1.4), and therefore (1.6), is stronger than needed to imply strong consistency of all MWE's. Weaker conditions will be considered in a later paper.

The results of this paper remain true if inner probability is replaced by outer probability in (1.1), (1.2), and (1.3).

2. The dominated and weakly dominated cases.

We first investigate the relations among Definitions 1-8 and the condition (recall (1.6))

$$(2.1) \qquad {}_*P[\limsup_{n\to\infty} \sup_\Gamma y_n(\theta) < 0] = 1.$$

Theorem 2.1. The following equivalences hold among Definitions 1-8: $1 \Leftrightarrow 5$, $2 \Leftrightarrow 6$, $3 \Leftrightarrow 7$, $4 \Leftrightarrow 8$. Each of 3, 4, 7, 8 implies (2.1). Therefore $B \Leftrightarrow L$, $B(\text{weak}) \Leftrightarrow L(\text{weak})$, and each of these four conditions implies (1.6) for all P_0, r, so each implies strong consistency of all AMWE's.

Proof: We first show that $3 \Leftrightarrow 7$ (the proofs of the other three equivalences are similar). If y is dominated by 0 on Γ, let $j = k$ and $v(x_1, \ldots, x_j, \theta) = s(x_1, \ldots, x_k)$. Clearly v is a strongly measurable mapping into $B(\Gamma)$. Since s may be chosen such that $-\infty < Es < 0$, (v), (vi), and (vii) are satisfied, so $3 \Rightarrow 7$. Next suppose that y is Bochner-dominated by 0 on Γ. Letting

$$v_n(\theta) = \frac{1}{n} \sum_{i=0}^{n-1} v(x_{ij+1}, \ldots, x_{(i+1)j}, \theta)$$

it follows from the Strong Law of Large Numbers (SLLN) for Bochner integrable random variables taking values in a Banach space S that

$$P[\sup_\Gamma |v_n(\theta) - Ev(\theta)| \to 0] = 1.$$

(See Beck [1] or Hans [3]. In stating the SLLN they assume that the Banach space S is separable. However, even if this is not the case - e.g. if $S = B(\Gamma)$ - strong measurability implies that v is almost separably-valued ([4], p. 72) so the range of v lies in a separable closed linear subspace of S.) Therefore

$$P[\sup_\Gamma v_n(\theta) \to \sup_\Gamma Ev(\theta) < 0] = 1.$$

(By Criterion 4 of Hans [3], strong measurability implies that $v_n(\theta)$ is Borel measurable so $\sup_\Gamma v_n(\theta)$, being a continuous function of $v_n(\theta)$, is also Borel measurable, i.e., a random variable.) However, we may apply Lemma 2.7 (ii) with $y_n(\theta)$ replaced by $v_n(\theta)$, so there is an integer $m \geq 1$ such that $E\sup_\Gamma v_m(\theta) < 0$. Thus Definition 3 is satisfied if we take $k = mj$ and $s(x_1, \ldots, x_k) = \sup_\Gamma v_m(\theta)$.

We now show that $3 \Rightarrow (2.1)$ using an idea of Berk [2]. For any $n \geq k$, let $\alpha = \{\alpha_1, \ldots, \alpha_k\}$ denote a selection of k indices from $\{1, 2, \ldots, n\}$. Then

$$y_n(\theta) = \binom{n}{k}^{-1} \sum_\alpha [k^{-1} \sum_{i \in \alpha} y(x_i, \theta)]$$

so that for all $x_1, \ldots, x_n$ in a set of probability one

$$\sup_\Gamma y_n(\theta) \leq \binom{n}{k}^{-1} \sum_\alpha s(x_{\alpha_1}, \ldots, x_{\alpha_k}) \equiv S_{n,k}$$

(We choose s to be a symmetric function of $x_1, \ldots, x_k$). Berk ([2], p. 55-56) shows that $\{S_{n,k}\}_{n=1}^\infty$ forms a reverse martingale sequence and $S_{n,p} \to Es < 0$ almost surely as $n \to \infty$, which implies (2.1).

Finally we show that $4 \Rightarrow (2.1)$. If y is weakly dominated by 0 on Γ, there is a function $b(\theta)$ such that (applying the above argument)

$$(2.2) \qquad {}_*P[\limsup_{n\to\infty} \sup_\Gamma (y_n(\theta)/b(\theta)) < 0] = 1.$$

Since $\inf_{\Gamma} b(\theta) > 0$, this implies (2.1). □

Example 3. $X_1, X_2, \ldots$ are i.i.d. random variables, each with the normal distribution $N(-1, 1)$. Take $\Gamma = [1, \infty)$ and $y(x, \theta) = \theta x$, so $y_n(\theta) = \theta \bar{x}_n$. Since $\bar{X}_n \to -1$ a.s., $\sup_{\Gamma} y_n(\theta) \to -1$ a.s. so (2.1) is satisfied. However, for all n

$$P[\sup_{\Gamma} y_n(\theta) = \infty] = P[\bar{X}_n > 0] > 0$$

so y is not dominated on Γ. Choosing $b(\theta) = \theta$ we see that y is weakly dominated by 0 on Γ. Thus neither (2.1) nor Definition 4 necessarily implies that Definition 3 is satisfied.

A partial converse to Theorem 2.1 is presented in Lemma 2.7 (ii) and Theorem 2.8. Several preliminary results are needed, some of which are also applied in section 3.

Lemma 2.2. If X and Y are independent real-valued random variables, then $E(X + Y)^+ < \infty \Rightarrow E\, X^+ < \infty$ and $E\, Y^+ < \infty$.

Proof. Since $E(X + Y)^+ = E\{E[(X+Y)^+ | Y]\}$ it follows that $E(X + y)^+ < \infty$ for a.e. y. But $X^+ \le (X + y)^+ + |y|$ so $E\, X^+ < \infty$, and similarly $E\, Y^+ < \infty$. □

Lemma 2.3. If y is dominated or weakly dominated on Γ then for every θ' in Γ, $E[y_1(\theta')]^+ < \infty$. Thus for every θ' and n $E\, y_n(\theta') = E\, y_1(\theta')$ is well-defined (possibly $= -\infty$) and $\sup_{\Gamma} E\, y_1(\theta) < \infty$.

Proof. If y is dominated on Γ, Definition 1 implies that

$$y_k(\theta') \le \sup_\Gamma y_k(\theta) \le s(x_1, \ldots, x_k)$$

so $E[y_k(\theta')]^+ \le E\, s^+ < \infty$. The result then follows from Lemma 2.2 (the weakly dominated case is treated similarly).$\square$

Lemma 2.4. Suppose that for every θ' in Γ, $E\, y_1(\theta')$ is well-defined (possibly $\pm\infty$). Then

$$_*P[\sup_\Gamma E\, y_1(\theta) \le \liminf_{n\to\infty} \sup_\Gamma y_n(\theta)] = 1.$$

Proof. For each θ' in Γ and all n, $y_n(\theta') \le \sup_\Gamma y_n(\theta)$. Letting $n \to \infty$ the result follows from the SLLN. $\square$

Remark. Lemma 2.4 implies that

$$_*P[\sup_\Gamma E\, y_1(\theta) \le \limsup_{n\to\infty} \sup_\Gamma y_n(\theta)] = 1, \tag{2.3}$$

and that

$$_*P[\exists \lim_{n\to\infty} \sup_\Gamma y_n(\theta) = \sup_\Gamma E\, y_1(\theta)] = 1 \tag{2.4}$$

if and only if equality holds in (2.3), in fact, if and only if

$$_*P[\sup_\Gamma y_1(\theta) \ge \limsup_{n\to\infty} \sup_\Gamma y_n(\theta)] = 1. \tag{2.5}$$

Lemma 2.5. (i) If $E\, y_1(\theta') > -\infty$ for some θ' in Γ, then $E[\sup_\Gamma y_n(\theta)]^- < \infty$ for all n such that $\sup_\Gamma y_n(\theta)$ is measurable, so $E \sup_\Gamma y_n(\theta)$ is well-defined (possibly $= +\infty$).

(ii) If y is dominated on Γ, then $E[\sup_\Gamma y_n(\theta)]^+ < \infty$ for all $n \ge k$ such that $\sup_\Gamma y_n(\theta)$ is measurable, so $E \sup_\Gamma y_n(\theta)$ is well-defined (possibly $= -\infty$).

<u>Proof</u>. (i) is trivial; (ii) follows from

$$\text{(2.6)} \qquad \sup_\Gamma y_n(\theta) \le \binom{n}{k}^{-1} \sum_\alpha \sup_\Gamma \left[k^{-1} \sum_{i\varepsilon\alpha} y(x_i, \theta)\right] \equiv Y_{n,k}(\Gamma)$$

(see the proof of Theorem 2.1). The result need not hold if y is weakly dominated on Γ, unless $b(\theta)$ is bounded above. □

<u>Lemma 2.6</u>. If $E\, y_1(\theta')$ is well-defined for all θ' in Γ and if $\sup_\Gamma y_n(\theta)$ is measurable and $E \sup_\Gamma y_n(\theta)$ is well-defined for all (sufficiently large) n, then

$$\text{(2.7)} \qquad \sup_\Gamma E\, y_1(\theta) \le \downarrow \lim_{n\to\infty} E \sup_\Gamma y_n(\theta).$$

<u>Proof</u>. From (2.6) with n, k replaced by $n + 1$, n,

$$\text{(2.8)} \qquad \sup_\Gamma E\, y_1(\theta) = \sup_\Gamma E\, y_{n+1}(\theta) \le E \sup_\Gamma y_{n+1}(\theta) \le E \sup_\Gamma y_n(\theta),$$

which implies (2.7). □

Under the hypothesis of Lemma 2.6, (2.3) is valid if $_*P$ is replaced by P, and should be compared with (2.7). The relationship between (2.3) and (2.7) is now clarified.

<u>Lemma 2.7 (i)</u>. If $\sup_\Gamma y_n(\theta)$ is measurable and $E \sup_\Gamma y_n(\theta)$ well-defined for all (sufficiently large) n then

$$\text{(2.9)} \qquad P\left[\limsup_{n\to\infty} \sup_\Gamma y_n(\theta) \le \lim_{n\to\infty} E \sup_\Gamma y_n(\theta)\right] = 1.$$

(ii) If y is dominated on Γ and $\sup_\Gamma y_n(\theta)$ is measurable for all (sufficiently large) n then

$$(2.10) \qquad P[\limsup_{n\to\infty} \sup_\Gamma y_n(\theta) = \lim_{n\to\infty} E \sup_\Gamma y_n(\theta)] = 1.$$

Therefore under this measurability assumption, y is dominated by 0 on Γ <u>if</u> <u>and</u> <u>only</u> <u>if</u> (2.1) is satisfied and y is dominated on Γ.

<u>Proof. (i)</u> Assume that the right-hand side of the inequality is $< \infty$ (in which case y is dominated on Γ; otherwise (2.9) is trivial). Referring to (2.6), for any q such that $E \sup_\Gamma y_q(\theta) < \infty$, $\{Y_{n,q}\}$ is a reverse martingale $(n=1, 2, \ldots)$ and

$$(2.11) \qquad Y_{n,q} \to E \sup_\Gamma y_q(\theta) \quad \text{a.s.} \quad \text{as} \quad n \to \infty,$$

so $\limsup_{n\to\infty} \sup_\Gamma y_n(\theta) \le E \sup_\Gamma y_q(\theta)$ a.s. Letting $q \to \infty$ we obtain (2.9).

(ii) We use (2.6) and (2.11) to apply a well-known extension of the Fatou-Lebesgue Theorem ([8], p. 162), obtaining

$$\lim_{n\to\infty} E \sup_\Gamma y_n(\theta) \le E[\limsup_{n\to\infty} \sup_\Gamma y_n(\theta)].$$

Combining this with (2.9) yields (2.10). □

Lemma 2.7(ii) implies that $\limsup \sup_\Gamma y_n(\theta)$ is a constant a.s. (this is proved under weaker restrictions in Lemma 3.4) so (2.1) is equivalent to

$$P[\limsup_{n\to\infty} \sup_{\Gamma} y_n(\theta) < 0] > 0.$$

Returning to the problem of strong consistency of AMWE's, consider the following conditions:

Condition D (D(weak)). For every P_0 in $\mathcal{P}$ and $r \geq 1$, $u(x, \theta; \theta_0)$ is dominated (weakly dominated) on $\Omega_r(P_0)$.

Theorem 2.1, Lemma 2.7(ii), and the subsequent remark immediately yield

Theorem 2.8. Suppose that for every θ_0, r, and (sufficiently large) n, $\sup_{\Omega_r} u_n(\theta; \theta_0)$ is measurable. Then the following are equivalent:

(a) Condition D and $P_0[\limsup_{n\to\infty} \sup_{\Omega_r} u_n(\theta; \theta_0) < 0] > 0$ (for all P_0, r)

(b) Condition D and (1.6) (for all P_0, r)

(c) Condition D and $P_0[\limsup_{n\to\infty} \sup_{\Omega_r} u_n(\theta; \theta_0) = c] = 1$ (for all P_0, r) where c is a constant (depending on P_0, r) such that $-\infty \leq c < 0$.

(d) Condition B (equivalently, Condition L).

Clearly B (weak) implies (1.6) (for all P_0, r) and D (weak). It is not known if, under suitable measurability assumptions, the converse is true. Note that (2.2) $\Rightarrow$ (2.1) but not necessarily conversely.

We now anticipate section 3 by presenting conditions under which equality holds in (2.3). This lemma is a generalization of a result due in essence to Wald [9], which was first stated formally by LeCam ([7], p. 300; see also Theorem 1 of Huber [5]). An extension of this lemma is given in Lemma 3.2 (ii).

Lemma 2.9. Let Γ be a compact Hausdorff space such that y is dominated on Γ. For every θ' in Γ suppose that

(i) there exists a decreasing sequence $\{G_m\} = \{G_m(\theta')\}$ of subsets of Γ, each a neighborhood of θ', such that for all m $\sup_{G_m} y_k(\theta)$ is measurable (k as in Definition 1), and

(ii) for all $x_1,\cdots, x_k$ in a set of probability 1 (which may depend on θ')

$$\sup_{G_m} y_k(\theta) \downarrow y_k(\theta') \quad \text{as} \quad m \to \infty.$$

Then equality holds in (2.3), implying (2.4).

Proof. By (i), (ii), dominance, the Monotone Convergence Theorem, and Lemmas 2.3 and 2.5, for each θ'

$$E \sup_{G_m} y_k(\theta) \downarrow Ey_1(\theta') \quad \text{as} \quad m \to \infty.$$

Thus, given $\delta > 0$ there is an integer $\mu = \mu(\theta')$ such that

$$E \sup_{G_\mu} y_k(\theta) \le \sup_\Gamma Ey_1(\theta) + \delta. \tag{2.12}$$

There is a finite subset $\{\theta_1',\cdots, \theta_h'\}$ of Γ such that $F_1,\cdots, F_h$ covers Γ, where $F_i = G_{\mu(\theta_i')}$. By (2.6), for $n \geq k$ and each $i = 1,\cdots, h$

$$\sup_{F_i} y_n(\theta) \le Y_{n,k}(F_i) \to E \sup_{F_i} y_k(\theta) \quad \text{a.s.}$$

as $n \to \infty$, so by (2.12)

$$_*P[\limsup_{n\to\infty} \sup_{F_i} y_n(\theta) \le \sup_\Gamma Ey_1(\theta)] = 1. \tag{2.13}$$

This, combined with the fact that

$$\limsup_{n\to\infty} \sup_\Gamma y_n(\theta) = \limsup_{n\to\infty} \max_i \sup_{F_i} y_n(\theta),$$

implies (2.5), completing the proof. □

Condition (ii) is satisfied if $\cap G_m = \{\theta'\}$ and $y(x, \cdot)$ is upper semicontinuous at θ' except for x in a P-null set (possibly depending on θ'). The measurability assumption in (i) is satisfied (for example) if Γ is a separable space, each G_m is open in Γ, and $y(x, \cdot)$ is lower semicontinuous on Γ for a.e. x.(For common spaces Γ other criteria for measurability may be more useful, such as right continuity if θ is a real-valued parameter.) In particular if Γ is separable, $\cap G_m = \{\theta'\}$, G_m open in Γ, and $y(x, \cdot)$ is continuous on Γ for a.e. x then (i) and (ii) are satisfied.

Returning to the problem of consistency of AMWE's, make the following additional assumptions: (α) Θ is locally compact, so we can choose $\{V_r(\theta_0)\}$ to be a sequence of <u>compact</u> neighborhoods of θ_0; (β) for every P_0 and r, $\sup_{V_r} E_0 u_1(\theta; \theta_0) = 0$; ($\gamma$) for every P_0 and r the conditions of Lemma 2.9 are satisfied for $(y, \Gamma, P) = (u(x, \theta; \theta_0), V_r, P_0)$. Then (1.4) $\Rightarrow$ (1.6), so these are equivalent conditions. If in addition Condition D and the measurability assumption of Theorem 2.8 are satisfied then by Lemma 1.1, Theorem 2.8 and the equivalence of (1.4) and (1.6) it follows that Condition B is <u>necessary</u> and <u>sufficient</u> for the strong consistency of all AMWE's.

In most cases of interest (α), (β), and (γ) are satisfied but not necessarily Condition D. In Example 4 (see also Example 5) we present a case where (α), (β),(γ) hold, where Condition B (weak) is satisfied so all AMWE's are strongly consistent, but where Condition D (hence B) fails. Thus the assumption of dominance (D) is too restrictive in general. Condition B (weak), requiring only weak dominance, has the disadvantage that we must look for suitable functions $b(\theta)$ (see the discussion after

Example 4). Therefore another condition (U), requiring neither dominance nor weak dominance, is introduced in section 3.

Before presenting Example 4 we show that Huber's assumptions A2-5 imply B (weak). Similarly it can be shown that the assumptions of [6], [9], and [10] (p. 320) imply B. A lemma is needed first.

Lemma 2.10. (i) y dominated on Γ and $\Gamma' \subset \Gamma \Rightarrow y$ dominated on Γ'. This remains true if "dominated" is replaced by "weakly dominated," "dominated by 0," or "weakly dominated by 0." (ii) y dominated on Γ_i, $i = 1, \ldots, h \Rightarrow y$ dominated on $\cup \Gamma_i$. This remains true if "dominated" is changed as above.

Proof. (i) is trivial. We prove (ii) for $h = 2$ with "dominated" replaced by "dominated by 0," as this is the more difficult case. Using (i) if necessary we may assume that Γ_1 and Γ_2 are disjoint. Let $\Gamma = \Gamma_1 \cup \Gamma_2$ and let k_i and $s_i = s_i(x_1, \cdots, x_{k_i})$ be as in Definitions 3 and 1, $i = 1, 2$. Let $m = k_1 k_2$,

$$S(i) = \frac{1}{k(i)} \sum_{j=0}^{k(i)-1} s_i(x_{jk_i+1}, \cdots, x_{(j+1)k_i})$$

where $k(1) = k_2$, $k(2) = k_1$, and

$$y^*(\theta) = \begin{cases} S(1) & \text{if } \theta \in \Gamma_1 \\ S(2) & \text{if } \theta \in \Gamma_2 \end{cases}$$

so $y_m(\theta) \leq y^*(\theta)$ and $y_{mn}(\theta) \leq y_n^*(\theta)$. Since $ES(i) = Es_i < 0$,

$$\sup_\Gamma y_n^*(\theta) = \max(S_n(1), S_n(2)) \xrightarrow{a.s.} \max(Es_1, Es_2) < 0.$$

Applying Lemma 2.7 (ii) with y replaced by y^*, this implies that $E \sup_\Gamma y_p^*(\theta) < 0$ for some p. Setting $k = mp$ and $s(x_1, \cdots, x_k) = \sup_\Gamma y_p^*(\theta)$, we conclude that y is dominated by 0 on Γ. Note that we cannot set $k = m$ and $s = \sup_\Gamma y^*(\theta) = \max(S(1), S(2))$, since $E\, S(i) < 0$ need not imply that $E \max(S(1), S(2)) < 0$. Also, the above proof is actually easier if one used the equivalence of Definitions 3 and 7.□

<u>Proposition 2.11</u>. Assumptions A2-5 of Huber [5] imply condition B (weak).

<u>Proof</u>. Huber's $\rho(x, \theta)$ corresponds to $-u(x, \theta)$ and his $\gamma(\theta_0)$ is zero in our case (see the remark following (5) of [5]) so infima rather than suprema appear in his paper. He takes V_r to be an open neighborhood U of θ_0 (see Theorem 1 of [5]) so $\Omega_r' = U'$ (prime denotes complement). In Lemma 1 of [5] (see (7)) it is shown that there exists

a compact set C such that $u(x, \theta)$ is weakly dominated by 0 on C'. In Theorem 1 of [5] (see (11)) it is shown that $C - U$ can be covered by a finite union of open sets U_s, $s = 1, \cdots, N$, such that $u(x, \theta)$ is dominated by 0 on each U_s, so $u(x, \theta)$ is dominated by 0 on $C - U$ (Lemma 2.10). Since $\Omega_r = U' \subseteq C' \cup (C - U)$, it follows from Lemma 2.10 that $u(x, \theta)$ is weakly dominated by 0 on Ω_r, implying B (weak). This result remains true if in A-5, (i) is replaced by weak dominance and any "suitable compactification" ([10], p.320) is used. □

Example 4. (Consistent estimation of the population median; see [5], p.223). Using the notation of Example 2 and assuming without loss of generality that $\theta_0 = 0$, we have $u(x, \theta) = |x| - |x - \theta|$. With $\Omega_r = (-\infty, -r^{-1}] \cup [r^{-1}, \infty)$ and $b(\theta) = |\theta|$ (Huber takes $b(\theta) = |\theta| + 1$ but this is not necessary) it is easy to verify that Huber's A1-5 hold, implying B (weak), so (1.6) holds (for all P_0, r) and hence all AMWE's are strongly consistent. It would be quite difficult to verify B (weak) directly since the (smallest) integer k such that $E \sup_{\Omega_r} [u_k(\theta)/b(\theta)] < 0$ obviously depends on r (in fact $k = k(r)$ increases to ∞ as $r \to \infty$) and is difficult to determine. Note also that in this example condition B fails: for all r and k

$$\sup_{\Omega_r} u_k(\theta) \geq h_r(x_1, \cdots, x_k) \equiv \begin{cases} k^{-1}\min|x_i| & \text{all } x_i \geq r^{-1} \\ -r^{-1} & \text{otherwise} \end{cases}$$

and $Eh_r(X_1, \cdots, X_k) = \infty$ if $X_1, X_2, \cdots$ are i.i.d., each distributed on $(-\infty, \infty)$ symmetrically about 0 and each $|X_i|$ having c.d.f.

$$F(x) = \frac{\log(1+x)}{1+\log(1+x)} \qquad (x \geq 0).$$

Thus in general B (weak) $\not\Rightarrow$ B and y weakly dominated by 0 on $\Gamma \not\Rightarrow y$ dominated on Γ.

We conclude this section with several remarks concerning the function $b(\theta)$ in Definition 4. If $y(x, \theta)$ is not itself dominated by 0 on Γ no general conditions guaranteeing the existence of such a function $b(\theta)$ are known, and if such a function does exist no general formula for $b(\theta)$ is known. Necessary conditions for the existence of $b(\theta)$ are that $\sup_\Gamma Ey_1(\theta) < 0$ and $b(\theta) \leq \delta|Ey_1(\theta)|$ for some $\delta > 0$ (by (2.1) and Lemma 2.4). This suggests choosing $b(\theta) = |Ey_1(\theta)|$ (as in Example 3) or more generally choosing $b(\theta)$ such that $b(\theta)/|Ey_1(\theta)|$ is bounded away from 0 and ∞ (as in Example 4 - see also Example 5). This "rule of thumb" seems to be satisfactory in most statistical applications but, at the level of generality of this paper, is not universally valid. To see this we now present an example where y is not dominated by 0 but is weakly dominated by 0, and where $b(\theta)$ cannot be $|Ey_1(\theta)|$. Let $\Gamma = [1, \infty)$ and let W be a random variable assuming the values 2 and -2 with probabilities 1/4 and 3/4, respectively. Let U be a stochastic process with parameter space Γ, U independent of W, such that for each θ $U(\theta)$ is uniformly distributed on the interval $(-2\theta^2, -\theta)$ and $\{U(\theta): \theta \in \Gamma\}$ is a set of mutually independent random variables, i.e., U is a "white noise" process. Let $V(\theta) = \theta W$ and let $X = \{X(\theta)\} \equiv \{U(\theta) + V(\theta)\}$. Let $X_1, X_2, \cdots$ be a sequence of i.i.d. stochastic processes, each having the same distribution as X. Let $\mathcal{X}$ be the set of all real valued functions on Γ, and for $(x, \theta) \in \mathcal{X} \times \Gamma$ set $y(x, \theta) = x(\theta)$. Then $P[\sup_\Gamma y_k(\theta) = \infty] > 0$ for every k, so y is not dominated. Setting $b(\theta) = \theta$, $y(x, \theta)/\theta = W + U(\theta)/\theta$ and $\sup_\Gamma[y_1(\theta)/\theta] \leq W - 1$, so y is weakly dominated by 0. However, $2|Ey_1(\theta)| = 3\theta + \theta^2$ and for every k, $\sup_\Gamma[y_k(\theta)/(3\theta + \theta^2)] \geq 0$ with probability 1.

This example shows that the choice of $b(\theta)$, if it exists, is very delicate: if $b(\theta)$ is too small, $y(x, \theta)/b(\theta)$ may not even be dominated, while if $b(\theta)$ is too large $y(x, \theta)/b(\theta)$ may be dominated but not dominated by 0. Also, there are situations where no such $b(\theta)$ exists: let $\Gamma[1, \infty)$ and let X be a stochastic process with parameter space Γ such that $\{X(\theta): \theta \in \Gamma\}$ are mutually independent and each $X(\theta)$ is uniformly distributed on $(-1, -1/\theta)$. With $y(x, \theta)$ as defined in the preceding paragraph, $y(x, \theta)$ is dominated on Γ but not weakly dominated by 0 on Γ. Note that if $b(\theta) = 1/\theta$, $y(x, \theta)/b(\theta)$ is dominated by 0, but $\inf b(\theta) = 0$. If the example is changed slightly so that $X(\theta)$ is uniformly distributed on $(-1, 0)$, then there is no function $b(\theta) > 0$ such that $y(x, \theta)/b(\theta)$ is dominated by 0, even if we do not require that $\inf b(\theta) > 0$.

It should be clear by now that the conditions presented in this section, due to the attempt to achieve wide generality, do not eliminate the usefulness of conditions such as Huber's which may be easier to verify in applications. Perhaps the most important result of this section is Lemma 2.7 (ii) which presents, under the regularity assumptions of dominance and measurability, a <u>necessary</u> <u>and</u> <u>sufficient</u> condition for (2.1)(and (1.6)), namely

$$(2.14) \qquad \downarrow\lim_{n\to\infty} E \sup_{\Gamma} y_n(\theta) < 0.$$

It is not always easy to verify this condition if it is in fact satisfied (see Examples 4, 6), but if it is in fact false this is usually easy to recognize. In the next section we see that (3.1) and (3.6) are, under some regularity assumptions that do not seem very restrictive, also necessary and sufficient for (2.1) and are often easier to verify (see Lemma 3.3 (ii) and the subsequent remark.)

3. A new condition for strong consistency.

Throughout this section it is assumed that for every θ in Γ (respectively Θ), $Ey_1(\theta)$ is well-defined (respectively $Eu_1(\theta)$). We are seeking conditions which imply (2.1) (resp. (1.6)); one such condition is that y be (weakly) dominated by 0 on Γ. A different approach is suggested by Lemmas 2.4 and 2.9. A necessary condition for (2.1) to hold is that

$$\sup_\Gamma Ey_1(\theta) < 0, \tag{3.1}$$

and if equality holds in (2.3), as in Lemma 2.9, then (3.1) is necessary and sufficient for (2.1). Lemma 2.9 is too restrictive for our purposes, however, since it requires that Γ (and Ω_r) be compact which is not the case in many statistical applications. We therefore present an extension (Lemma 3.2 (ii) and subsequent remark) which requires only that Γ be σ-compact, i.e., a countable union of compact sets.

Lemma 3.1. Let $\{\beta(n, h)\}$ be a double sequence of extended real numbers $(-\infty \leq \beta(n, h) \leq \infty)$ such that for each n, $\beta(n, h)$ increases to a limit $\beta(n, \infty)$.

(i) Let

$$\beta(h) = \limsup_{n\to\infty} \beta(n, h) \tag{3.2}$$

so $\beta(h)$ increases to some limit $\beta(\infty)$ $(-\infty \leq \beta(\infty) \leq \infty)$. Then

$$\beta(n, \infty) < \infty \quad \text{(fasln)} \quad \text{and} \tag{3.3}$$

$$\beta(n, h) \uparrow \beta(n, \infty) \text{ uniformly in } n \text{ as } h \to \infty \tag{3.4}$$

together imply that $\limsup_{n\to\infty} \beta(n, \infty) = \beta(\infty)$.

(ii) Suppose that for each h, the limit $\beta(h) = \lim_{n\to\infty} \beta(n, h)$ exists (possibly infinite), so $\beta(h)$ increases to some limit $\beta(\infty)(-\infty \leq \beta(\infty) \leq \infty)$. Then (3.3) and (3.4) together imply that

$$(3.5) \qquad \lim_{n\to\infty} \beta(n, \infty) = \beta(\infty).$$

If in addition $-\infty < \beta(\infty) < \infty$ then the converse is also true, i.e., (3.5) implies (3.3) and (3.4).

Proof. (i) Since $\beta(n, \infty) \geq \beta(n, h)$, $\limsup \beta(n, \infty) \geq \beta(h)$ for all h, so $\limsup \beta(n, \infty) \geq \beta(\infty)$. To prove the opposite inequality let $\{n_\nu\}$ denote those values of n such that $-\infty < \beta(n_\nu, \infty)$ (if there are only finitely many such n, the result is trivally true) so $\limsup \beta(n, \infty) = \limsup \beta(n_\nu, \infty)$ and $\beta(h) = \limsup \beta(n_\nu, h)$ By (3.3) and (3.4) there exist N and H such that $n_\nu \geq N$ and $h \geq H$ imply that $\beta(n_\nu, h)$ and $\beta(n_\nu, \infty)$ are finite. Then the desired inequality follows from (3.4) and the identity

$$\beta(n_\nu, \infty) = [\beta(n_\nu, \infty) - \beta(n_\nu, h)] + \beta(n_\nu, h).$$

(ii) First assume (3.3) and (3.4). Since $\beta(n, \infty) \geq \beta(n, h)$, $\liminf \beta(n, \infty) \geq \beta(h)$ for all h so $\liminf \beta(n, \infty) \geq \beta(\infty) = \limsup \beta(n, \infty)$ (by (i)), proving (3.5). Next assuming (3.5) and $-\infty < \beta(\infty) < \infty$, (3.3) is obviously satisfied. Given $\delta > 0$ choose H such that $\beta(\infty) \leq \beta(H) + \delta/3$ and choose N such that for $n \geq N$, $|\beta(n, H) - \beta(H)| \leq \delta/3$ and $|\beta(n, \infty) - \beta(\infty)| \leq \delta/3$. Then if $h \geq H$ and $n \geq N$,

$$\beta(n, \infty) - \beta(n, h) \leq \beta(n, \infty) - \beta(n, H) \leq \delta,$$

which implies (3.4).□

<u>Lemma 3.2</u>. Suppose that $\Gamma = \cup_{h=1}^{\infty}\Gamma_h$ where $\Gamma_h \subset \Gamma_{h+1}$ for all h.

(i) The two conditions

$$(3.6) \qquad {}_*P[\sup_\Gamma y_n(\theta) < \infty \;\; \text{fasln}] = 1 \quad \text{and}$$

$$(3.7) \qquad {}_*P[\sup_{\Gamma_h} y_n(\theta) \uparrow \sup_\Gamma y_n(\theta) \;\; \text{uniformly in } n \text{ as } h \to \infty] = 1$$

together imply that

$$(3.8) \qquad {}_*P[\limsup_{n\to\infty} \sup_\Gamma y_n(\theta) = \lim_{h\to\infty} \limsup_{n\to\infty} \sup_{\Gamma_h} y_n(\theta)] = 1.$$

(ii) Suppose that for each h

$$(3.9) \qquad {}_*P[\limsup_{n\to\infty} \sup_{\Gamma_h} y_n(\theta) = \sup_{\Gamma_h} Ey_1(\theta)] = 1.$$

Then (3.6) and (3.7) together imply that

$$(3.10) \qquad {}_*P[\limsup_{n\to\infty} \sup_\Gamma y_n(\theta) = \sup_\Gamma Ey_1(\theta)] = 1,$$

i.e., equality holds in (2.3). If in addition $-\infty < \sup_\Gamma Ey_1(\theta) < \infty$, then (3.10) implies (3.6) and (3.7).

<u>Proof</u>. If we set $\beta(n, h) = \sup_{\Gamma_h} y_n(\theta)$, (i) follows from Lemma 3.1 (i). In (ii), (3.9) implies that the limit $\beta(h) = \lim_{n\to\infty} \beta(n, h) = \sup_{\Gamma_h} Ey_1(\theta)$ exists (recall (2.4)) so the result follows from Lemma 3.1(ii).□

Often we can take each Γ_h to be compact so Γ is σ - compact. Then (3.9) will be satisfied, for example, if for every h, (y, Γ_h) satisfy the conditions of Lemma 2.9 (see Examples 5, 6).

<u>Lemma 3.3</u>. Suppose that $\Gamma = \cup_{h=1}^{\infty}\Gamma_h$ where $\Gamma_h \subset \Gamma_{h+1}$.

(i) If (3.7) is satisfied then (2.1) $\Leftrightarrow$ (3.6) and

(3.11) $$_{*}P[\lim_{h\to\infty} \limsup_{n\to\infty} \sup_{\Gamma_h} y_n(\theta) < 0] = 1.$$

(ii) If (3.7) and (3.9) hold then (2.1) $\Leftrightarrow$ (3.6) and (3.1). (Note that (2.1) $\Rightarrow$ [(3.6) and (3.11)] $\Rightarrow$ [(3.6) and (3.1)] in general.)

Proof. Immediate, by Lemma 3.2.□

If each Γ_h is compact, condition (3.9) usually may be verified using Lemma 2.9 without much difficulty, so the critical regularity assumption needed to apply Lemma 3.3(ii) is (3.7). Once this can be verified it is relatively simple to evaluate the supremum in (3.1) (whereas it might be much more troublesome to evaluate the expression in (2.14) - compare Examples 4 and 5).

Part (i) of Lemma 3.3 represents a more general method than part (ii) since (3.9) is not needed. However, the author knows of no statistical problems where (i) is applicable but (ii) is not.

Before returning to the problem of consistency of estimates we present a result which was alluded to following Lemma 2.7.

Lemma 3.4. Suppose that (3.6) is satisfied and that $\sup_\Gamma y_n(\theta)$ is measurable for all (sufficiently large) n. Then there exists a constant c, $-\infty \leq c \leq \infty$, such that

$$P[\limsup_{n\to\infty} \sup_\Gamma y_n(\theta) = c] = 1.$$

(By Lemma 2.4, $c \geq \sup_\Gamma Ey_1(\theta)$.)

Proof. Let $A_n = \{\sup_\Gamma y_n(\theta) < \infty\}$. For fixed N and all $n > N$

$$\sup_\Gamma y_n(\theta) \leq \frac{N}{n} \sup_\Gamma y_N(\theta) + \frac{n-N}{n} \sup_\Gamma (n-N)^{-1}[y(x_{N+1}, \theta) + \cdots + y(x_n, \theta)].$$

Thus, on A_N

$$(3.12) \qquad Z \equiv \limsup_{n\to\infty} \sup_\Gamma y_n(\theta) \leq \limsup_{n\to\infty} \sup_\Gamma (n-N)^{-1}[y(X_{N+1}, \theta) + \cdots + y(X_n, \theta)] \equiv Y_N.$$

Note that Z and Y_N are identically distributed and that Y_N depends only on $X_{N+1}, X_{N+2}, \cdots$. Thus for any real number c, $\limsup\{Y_k \leq c\} = \{Y_k \leq c$ occurs infinitely often$\}$ is a tail event and hence has probability either 0 or 1 by the Kolmogorov Zero-One Law. Now $P[\liminf A_n] = 1$ by (3.6) so

$$\begin{aligned} P[Z \leq c] &= P[\{Z \leq c\} \cap \{\liminf A_n\}] \\ &\geq P[\{\limsup\{Y_k \leq c\}\} \cap \{\liminf A_n\}] \\ &= P[\limsup\{Y_k \leq c\}] \\ &\geq \lim_{k\to\infty} P[Y_k \leq c] \\ &= P[Z \leq c], \end{aligned}$$

where the first inequality follows from (3.12). Thus $P[Z \leq c] = P[\limsup\{Y_k \leq c\}] = 0$ or 1 for every c, so Z must be a degenerate random variable.□

If $\sup_\Gamma y_n(\theta)$ is measurable (fasln) this result implies that (2.1) ⇔ [(3.6) and (3.13)] ⇔ (3.14), where

(3.13) $\quad P[\limsup_{n\to\infty} \sup_\Gamma y_n(\theta) < 0] > 0$ and

(3.14) $\quad P[\limsup_{n\to\infty} \sup_\Gamma y_n(\theta) = c] = 1$ for some c, $-\infty \le c < 0$.

Furthermore, if in Lemmas 3.2 and 3.3 $\sup_{\Gamma_h} y_n(\theta)$ is measurable for all h and n, then ${}_*P$ may be replaced by P and "$= 1$" may be replaced by "> 0" in (3.9) and (3.11).

We now use these results to present a new condition for strong consistency of AMWE's, not requiring (weak) dominance.

Condition U. For every P_0 in $\mathcal{P}$ and every integer $r \ge 1$, the triple $(y(x, \theta), \Gamma, P) = (u(x, \theta; \theta_0), \Omega_r(\theta_0), P_0)$ satisfies (3.7), (3.9), (3.6), and (3.1) (for a suitable $\{\Gamma_h\}$).

We can now state the following theorem which is analogous to Theorems 2.1 and 2.8 in the dominated case.

Theorem 3.5. Condition U implies (1.6) for all P_0, r and therefore implies the strong consistency of all AMWE's. If for all P_0, r, and (sufficiently large) n, $\sup_{\Omega_r} u_n(\theta; \theta_0)$ is measurable and (3.7) and (3.9) are satisfied with $(y, \Gamma, P) = (u, \Omega_r, P_0)$, then the following are equivalent:

(a) (3.6) and (3.13) with $(y, \Gamma, P) = (u, \Omega_r, P_0)$ (for all P_0, r)

(b) (1.6) (for all P_0, r)

(c) (3.14) with $(y, \Gamma, P) = (u, \Omega_r, P_0)$ (for all P_0, r)

(d) (3.6) and (3.1) with $(y, \Gamma, P) = (u, \Omega_r, P_0)$ (for all P_0, r).

Measurability is not required for the equivalence of (b), (c) and (d) if ${}_*P$ is used in (3.14).

Now, as in the second paragraph after Lemma 2.9 make the additional assumptions (α), (β), (γ). If in addition we assume

Condition R. For every P_0 and r

(i) each $\Omega_r(\theta_0)$ is σ-compact,

(ii) with $(y, \Gamma, P) = (u(x, \theta; \theta_0), \Omega_r(\theta_0), P_0)$, (3.9) is satisfied.

(iii) with (y, Γ, P) as in (ii), (3.7) is satisfied.

(rather than Condition D as we did in section 2), then Condition U is necessary and sufficient for the strong consistency of all AMWE's. Note that if (γ) is satisfied, R(ii) usually will also be satisfied. The crucial condition is R(iii), just as dominance was crucial in section 2. However, it seems that R(iii) is less restrictive than dominance, as seen by the following Example 5. Even when both are satisfied, Condition U may be easier to verify than B; as illustrated in Example 6.

Example 5. (continuation of Example 4). We show that Condition U is satisfied. Fix P_0 and r, and assume $\theta_0 = 0$. With $\Gamma = \Omega_r$ as in Example 4, let $\Gamma_h = [-h, -r^{-1}] \cup [r^{-1}, h]$, a compact set. Since $u_1(\theta) \leq h$ on Γ_h and $u_1(\theta)$ is continuous, Lemma 2.9 can be applied to verify (3.9), and (3.6) is obviously true. To verify (3.7) note that $u_n(\theta)$ is a unimodal function with mode at X[(n+1)/2] (the (n+1)/2 - th order statistic from a sample of size n) if n is odd, and mode "plateau" on the interval (X[n/2], X[(n/2)+1]) if n is even. Thus, if we choose δ such that

$$P_0[-\delta \leq X_1 \leq \delta] > \frac{1}{2}$$

it follows that

$$P_0[-\delta \leq \text{mode (plateau) of } u_n(\theta) \leq \delta \quad \text{fasln}] = 1,$$

which implies (3.7). Thus it remains only to verify (3.1). If $\theta \geq 0$ ($\theta \leq 0$ is similar)

$$\begin{aligned} E_0\, u_1(\theta) &= 2\{\theta\, P_0[X_1 \geq \theta] + \int_{0+}^{\theta-} x\, d\, P_0(x)\} - \theta \\ &\leq \theta\{2P_0[X_1 > 0] - 1\} \\ &\leq 0. \end{aligned}$$

The first inequality is strict if $P_0[0 < X_1 < \theta] > 0$ and the second is strict if $P_0[X_1 > 0] < \frac{1}{2}$. Thus $P_0[X_1 > 0] < \frac{1}{2}$ implies $\sup_{\Omega_r} E_0\, u_1(\theta) < 0$. If $P_0[X_1 > 0] = \frac{1}{2}$ then for all

$\theta > 0$, $P_0[0 < X_1 < \theta] > 0$ since otherwise the population median would not be unique. This again implies $\sup_{\Omega_r} E_0\, u_1(\theta) < 0$ since this expectation decreases monotonically in $|\theta|$, so (3.1) holds and hence Condition U is satisfied.

In this example Huber's conditions perhaps are easier to verify, but Condition U did not require that we search for an unknown function $b(\theta)$ which seems to be a great advantage in general. Incidentally, $|E\, u_1(\theta)| \leq |\theta|$ since $|u_1(\theta)| \leq |\theta|$, which shows that if $b(\theta) = |\theta|$, then $b(\theta)/|Eu_1(\theta)|$ is bounded away from 0 (and ∞, as seen in section 2), thus verifying an earlier remark.

Example 6. In the situation of Example 1 let $\Theta = [0, \infty)$, μ = Lebesgue measure, and

$$f(x, \theta) = 1 + \frac{a\theta}{1+\theta} \sin(x - \theta) \qquad \text{if} \quad 0 \leq x \leq 2\pi,$$

$f(x, \theta) = 0$ otherwise, where $0 < a < 1$ is a constant. We wish to show that all approximate maximum likelihood estimators are strongly consistent, although neither Wald's, Kiefer and Wolfowitz's, nor Huber's conditions are satisfied here. Fix θ_0 and r, and let

$$\Omega_r = [0, \theta_0 - r^{-1}] \cup [\theta_0 + r^{-1}, \infty).$$

Since $u(x, \theta) = \log f(x, \theta) - \log f(x, \theta_0) \leq \log 2(1 + \theta_0)$, (3.6) is satisfied. In fact, u is dominated on Ω_r, but it would be very difficult to verify (2.14) directly. Setting $\Gamma = \Omega_r$ and $\Gamma_h = \Gamma \cap [0, h]$, Γ_h is compact, u is dominated on Γ_h, and u is continuous so Lemma 2.9 implies (3.9). To verify (3.7) note that $\log t$ is uniformly continuous on $[1 - a, 1 + a]$, that for any x_i

$$\left|\left[1 + \frac{a\theta}{1+\theta} \sin (x_i - \theta)\right] - \left[1 + a \sin (x_i - \theta)\right]\right| \leq \frac{1}{1+\theta} ,$$

and that both terms in square brackets lie in $[1 - a, 1 + a]$. Therefore given any $\delta > 0$ there exists $M > 0$ such that $\theta \geq M$ implies

$$\left| \frac{1}{n} \log \prod_{i=1}^{n} \left[1 + \frac{a\theta}{1+\theta} \sin (x_i - \theta)\right] - \frac{1}{n} \log \prod_{i=1}^{n} \left[1 + a \sin (x_i - \theta)\right]\right| \leq \delta$$

independently of n and $x_1, x_2, \cdots$. Thus for θ sufficiently large, $u_n(\theta)$ can be approximated arbitrarily closely by a periodic function, uniformly in n and $x_1, x_2, \cdots$, which implies (3.7). Finally (3.1) follows from the Information Inequality and an easy limiting argument, so Condition U is satisfied and all AMLE's are strongly consistent. Note that by Theorem 2.8 and dominance this implies Condition B, but as remarked above this would be difficult to demonstrate directly.

In Example 6, conditions (a), (b), (c) of Bahadur ([10], p.320) are satisfied, but this is not immediately evident. The difficulty arises when trying to find a "suitable compactification," for the obvious one-point compactification $[0, \infty]$ is not adequate. We must adjoin to $\Theta = [0, \infty)$ an entire interval of length 2π, say $I = [-2\pi, 0)$. Any θ in Θ can be uniquely represented as $\theta = 2\pi m + r$ where m is an integer and $-2\pi \leq r < 0$. Then the topology in $\bar{\Theta} = \Theta \cup I$ must be defined so that if $\{\theta_n\} \subset \Theta$ and $\bar{\theta} \in I$, $\theta_n \to \bar{\theta}$ if and only if $m_n \to \infty$ and $r_n \to \bar{\theta}$.

However, it has been pointed out to the author by Professor Bahadur that if in this example (or more generally in Example 1) we redefine $\theta(P) \equiv P$ and $\Theta \equiv \mathcal{P}$, and consider the topology of weak convergence in $\mathcal{P}$, then there is a natural compactification of $\mathcal{P}$, namely the closure of $\mathcal{P}$ in the set of all measures on $(\mathcal{X}, \mathcal{G})$ with total mass ≤ 1. This is in fact the natural parameterization and topology to consider in Example 1 since we are interested primarily in estimating the underlying probability distribution. Using this parameterization Bahadur's conditions (a), (b), (c) are easily verified in Example 6.

A similar situation occurs in the well-known special case of Example 1 where $\theta = (\mu, \sigma^2)$ and $f(x, \theta)$ is the density of the normal distribution $N(\mu, \sigma^2)$ (see [6], p.904). Here again both Conditions B and U hold, implying strong consistency of all AMLE's, and it is again difficult to find a "suitable compactification." If, however, the natural parameterization $\theta(P) = P$ and the natural compactification described above are considered, Bahadur's conditions (a) and (c) are easily verified. Condition (b) fails, but as pointed out by Kiefer and Wolfowitz (whose remark provided much of the motivation for this paper) it can be replaced by the assumption of dominance with $k = 2$.

We can also formulate Condition U(weak) by replacing the triple (u, Ω_r, P_0) in Condition U by $(u/b(\theta), \Omega_r, P_0)$ for a suitable function $b(\theta) = b(\theta;\ \theta_0, r)$ with $\inf b(\theta) > 0$. At this time, however, no examples are known where U fails but U(weak) is applicable. Also uncertain are some of the relations among B, B(weak), U, and U(weak). Using "white noise" as in section 2, it is easy to find (y, Γ, P) where y is dominated by 0 on Γ but where there does not exist a function

$b(\theta)$ (with $\inf_\Gamma b(\theta) > 0$) such that $(y/b(\theta), \Gamma, P)$ satisfies (3.9), which suggests that $B \not\Rightarrow U$ or U(weak). Also, U need not imply B (Examples 4, 5). It is not known, though, if U or U(weak) imply B(weak). It would also be interesting to know if Huber's conditions A1-5 imply U or U(weak) (the converse was shown to be false), and if Bahadur's conditions (a), (b), (c) imply Condition U. If we consider the natural parameterization and compactification described above it is suspected that the answer to this last question is yes, and that in fact many of the questions posed in this paper can be answered definitely in this context. This is the subject of current investigations.

The conditions introduced in this paper (B, B(weak), U, U(weak)) are all sufficient conditions for consistency of likelihood ratio tests.

Acknowledgements. I wish to thank Professors R. R. Bahadur, Herman Chernoff, Lucien LeCam, Naresh Jain, and William Sudderth for several helpful discussions and correspondences.

REFERENCES

[1] Beck, Anatole (1963). On the strong law of large numbers. _Ergodic Theory_ (edited by F. B. Wright) 21-53. Academic Press, New York.

[2] Berk, Robert H. (1966). Limiting behavior of posterior distributions when the model is incorrect. _Ann. Math. Statist._ 37, 51-58.

[3] Hans, O. (1956). Generalized random variables. _Trans. First Prague Conference Information Theory, etc._ 61-104. Publishing House of the Czechoslovak Academy of Sciences, Prague.

[4] Hille, E. and Phillips, R. S. (1957). _Functional Analysis and Semi-Groups._ (revised edition). Amer. Math. Soc. Colloquium Publication Vol. XXXI, Providence, Rhode Island.

[5] Huber, Peter J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. _Proc. Fifth Berkeley Symp. Math. Statist. and Probability_ Vol. I, 221-233. University of California Press, Berkeley.

[6] Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. _Ann. Math. Statist._ 27, 884-906.

[7] LeCam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. _Univ. Calif. Publications Statist._ 1, 277-328.

[8] Loève, M. (1963). _Probability Theory_ (3rd ed.). Van Nostrand, Princeton.

[9] Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. _Ann. Math. Statist._ 20, 595-601.

[10] Bahadur, R. (1967). Rates of convergence of estimates and test statistics. _Ann. Math. Statist._ 38, 303-324.